



NOTA DE TEMA

Inteligência Artificial: Treinamento de Máquinas e Direito Autoral

Como proteger os criadores sem inviabilizar a inovação

Sobre o Conselho Digital

O Conselho Digital é uma entidade brasileira, sem fins lucrativos ou afiliações políticas, que coordena, estuda e representa o ecossistema dos aplicativos de internet e toda a diversidade dos seus modelos de negócios.

Nossa organização acredita que a tecnologia, quando bem construída e utilizada, é uma porta para o futuro. Ela nos mantém conectados, potencializa habilidades, desenvolve novas oportunidades e pode mudar a vida das pessoas para melhor.

Partindo dessa premissa, atuamos através de estudos, eventos e atividades de advocacy em favor de políticas públicas e setoriais que fortaleçam uma internet livre, segura e responsável no Brasil e no mundo.

Defendemos políticas que respeitem a neutralidade tecnológica, a inovação e a diversidade de modelos de negócios; e que tenham como consequência:

- Usuários conscientes e com poder de escolha;
- Uma sociedade plural e próspera;
- Ambientes de negócio juridicamente seguros;
- Mercados abertos e dinâmicos; e
- Empresas responsáveis e competitivas.

Por fim, assumimos o compromisso de construir um ambiente harmonioso e produtivo entre nossos associados, assim como uma relação transparente e colaborativa com a sociedade e governo.



Diretor-Executivo

www.conselhodigital.org.br

Takeaways – Posição do Conselho Digital

- **Treinamento de máquinas é a base do funcionamento dos sistemas de inteligência artificial.**
- **Treinar sistemas de inteligência artificial não é copiar obras, mas aprender padrões estatísticos; o modelo final não armazena nem reproduz conteúdos.**
- **O texto do Projeto de Lei nº 2338 aprovado no Senado trata o acesso a obras disponíveis publicamente para fins de treinamento de IA como se fosse reprodução.**
- **Existem quatro modelos regulatórios para o tratamento do treinamento de máquina: uso justo, opt-in, opt out e permissão expressa incluída em lei.**
- **O fair use (ou uso justo) maximiza a inovação, mas reduz a escolha do detentor do direito autoral e pode levar a judicialização. O opt-in maximiza o controle, mas inviabiliza o treinamento de IA prejudicando a inovação. O opt-out é o ponto intermediário funcional, mas pode prejudicar modelos já treinados. Permissões estatutárias, como a vigente no Japão, promovem segurança jurídica para o desenvolvimento de modelos com dados plurais e representativos, apesar de serem mais difíceis de serem implementados.**
- **Para funcionar, o opt-out precisa ser legível por máquina (com soluções simples como o robots.txt, como as já utilizadas pelos mecanismos de busca), permitindo escala, negociação e preservação da diversidade linguística — inclusive do português brasileiro e não poderia ser aplicável a modelos já anteriormente treinados a fim de preservar a segurança jurídica.**

Posicionamento do Conselho Digital sobre o Treinamento de Sistemas de Inteligência Artificial e Regimes de Opt-Out

- **O foco dessa nota é o input:** Antes de discutir modelos regulatórios, é fundamental distinguir **duas etapas diferentes** no funcionamento de sistemas de inteligência artificial: o **input**, que corresponde ao uso de conteúdos para treinar o sistema, e o **output**, que corresponde aos resultados gerados pelo sistema após o treinamento. O debate tratado nesta nota diz respeito exclusivamente ao **input**, isto é, às regras para o uso de conteúdos no treinamento de sistemas de IA. **Isso não altera, relativiza ou enfraquece** a aplicação de direitos autorais, direito de imagem, proteção de dados pessoais ou outras responsabilidades jurídicas relacionadas ao **output**.



- **Contexto do debate e funcionamento técnico da IA:** O avanço recente dos sistemas de **inteligência artificial (IA)**, especialmente aqueles baseados em **aprendizado de máquina**, trouxe para o centro do debate público a questão de como esses sistemas podem ser treinados e quais conteúdos podem ser utilizados nesse processo. Trata-se de um debate legítimo, mas frequentemente prejudicado por **equivocos sobre o funcionamento técnico da IA** e por **tentativas de aplicar automaticamente modelos regulatórios pensados para outros tipos de uso de obras e dados**. Para o Conselho Digital, qualquer discussão séria sobre o treinamento de

sistemas de inteligência artificial deve começar por uma explicação clara do que esse treinamento realmente é, avançar para uma análise honesta dos modelos regulatórios possíveis e considerar seus impactos sociais, econômicos, culturais, linguísticos e estratégicos.

- **O que é – e o que não é – treinar uma IA:** Treinar um sistema de inteligência artificial **não significa ensinar um computador a guardar textos, imagens, músicas ou obras**. Treinar uma IA significa **expor um modelo computacional a grandes volumes de dados para que ele aprenda padrões gerais**. De forma semelhante ao aprendizado humano, uma pessoa aprende a falar e escrever por exposição repetida à língua, sem necessariamente memorizar livros ou frases específicas. O que permanece é uma compreensão abstrata de como a língua funciona. No caso da IA, esse aprendizado ocorre por meio de **cálculos estatísticos e ajustes matemáticos**, e não por armazenamento de conteúdos.
- **Funcionamento técnico do treinamento:** Durante o treinamento, um sistema de IA analisa grandes volumes de textos, imagens, sons ou outros dados, identifica padrões recorrentes e relações estatísticas, ajusta milhões ou bilhões de parâmetros numéricos internos e utiliza os dados apenas de forma temporária, descartando-os ao final do processo. **O resultado do treinamento, assim, não é um acervo de conteúdos**, mas **um modelo matemático composto por números que representam probabilidades**, e não obras específicas. Por isso, um sistema de IA treinado pode gerar novos textos ou imagens com base em padrões aprendidos e generalizar conhecimento, mas **não contém um banco de dados de obras, não permite recuperar conteúdos específicos usados no treinamento e não reproduz conteúdos de forma automática**. Treinamento de IA não é reprodução, armazenamento ou redistribuição.
- **O mito da reprodução invisível:** Muitas propostas regulatórias partem da ideia de que o treinamento de IA cria uma espécie de “reprodução invisível” das obras utilizadas. Essa ideia é **tecnicamente incorreta**. O treinamento envolve apenas a exposição temporária a conteúdos para extração de padrões gerais, sem que esses conteúdos permaneçam acessíveis ou utilizáveis depois.

Confundir treinamento com reprodução leva a soluções regulatórias desproporcionais, que não atacam o problema real, pois o tipo de uso feito por uma aplicação de IA é radicalmente diferente do tipo de aproveitamento que uma pessoa faria de uma mesma obra.

Para pensar: Após o treinamento, o tamanho de uma aplicação de IA não é muito maior do que o era anteriormente ao treinamento, pois nenhuma obra é armazenada dentro do algoritmo.

- **O verdadeiro desafio regulatório:** Quando se compreende corretamente o funcionamento do treinamento, fica claro que o desafio regulatório **não é a proteção contra uma suposta cópia de obras**, mas sim **a governança do acesso automatizado e massivo a conteúdos para fins de treinamento de sistemas de IA**. Esse tipo de acesso existe desde o início da internet e sempre exigiu soluções que funcionem em escala.
- **Modelos regulatórios em disputa:** Nesse contexto, três modelos regulatórios se apresentam. No **fair use** ou **uso justo**, o treinamento é permitido sem possibilidade de exclusão, o que garante máxima viabilidade técnica, forte estímulo à inovação e diversidade de modelos, mas oferece pouco espaço para a escolha dos titulares e dificulta negociação e compensação. Além disso, o ordenamento jurídico brasileiro atualmente não comporta essa possibilidade. No **opt-in**, o treinamento só é permitido mediante autorização prévia e expressa, o que maximiza o controle formal, mas é **tecnicamente inviável em larga escala**, favorece a concentração econômica, exclui universidades e pequenos desenvolvedores e reduz drasticamente a diversidade tecnológica. Já o **opt-out** estabelece que o treinamento é permitido como regra, garantindo aos titulares o direito de se excluir, sendo compatível com a escala do treinamento de IA, preservando a escolha e criando base concreta para negociação. Porém, dependendo do modo como for implementado, pode afetar negativamente modelos já treinados.

- **A possibilidade adicional, permissão expressa incluída em lei:**
Além dos três modelos regulatórios acima, uma quarta possibilidade seria a promulgação de um dispositivo legal específico regulamentando o treinamento de obras protegidas por direito autoral. A previsão normativa expressa quanto à legalidade do acesso e uso de conteúdos protegidos para o treinamento de modelos eliminaria dúvidas e promoveria segurança jurídica. Porém, a propositura de quaisquer mudanças à legislação brasileira é um processo árduo e essa tem que ser vista como uma solução de longo-prazo.
- **Opt-out como ponto de equilíbrio e de negociação:** O opt-out, se previsto de forma eficiente, pode ocupar um ponto intermediário entre o treinamento sem barreiras do fair use e o bloqueio sistêmico do opt-in.
 - Trata-se de um modelo em que a análise computacional de obras (text and data mining) é permitida como regra, mas os titulares podem exercer o direito de exclusão dessa utilização para conteúdos apontados.
 - Esse desenho reconhece simultaneamente a necessidade de treinar sistemas de IA em escala e a legitimidade da escolha dos titulares.
 - Além disso, o **opt-out não elimina mercados — ele os viabiliza**: quando há um direito efetivo de se opor ao uso para TDM e treinamento, os conteúdos excluídos tornam-se identificáveis, acordos coletivos e setoriais passam a ser estruturáveis e os custos de transação diminuem.
 - Sem opt-out, o sistema tende ao uso irrestrito de fato; com opt-in, o sistema deixa de funcionar; com opt-out, há governança e negociação.
- **Implementação técnica e robots.txt:** Para funcionar, o opt-out não pode depender de pedidos manuais, contratos individuais ou avisos genéricos. Ele precisa ser **automaticamente interpretável por sistemas de IA**. A própria internet já oferece um precedente

consolidado para isso: o **robots.txt**, que há décadas permite que titulares indiquem, de forma simples e legível por máquina, se mecanismos de busca podem acessar e indexar seus conteúdos. Trata-se de um modelo escalável, barato, neutro e amplamente compreendido, cuja lógica pode ser aplicada ao treinamento de sistemas de IA. Além disso, a implementação do opt-out teria de ser feita de modo a não afetar modelos já treinados, a fim de preservar a segurança jurídica dos detentores desse tipo de modelo.

- **Impactos sobre atualidade, língua e soberania:** Restringir excessivamente o treinamento de sistemas de IA com conteúdos atuais torna esses sistemas rapidamente defasados, menos úteis e menos seguros, afetando sua aplicação em educação, saúde, serviços públicos e pesquisa. Da mesma forma, impedir o treinamento com conteúdos em **português brasileiro** faz com que sistemas de IA compreendam mal a língua, ignorem variações regionais e culturais, reproduzam vieses estrangeiros e aumentem a dependência tecnológica do país, comprometendo a **soberania cultural, informacional e tecnológica do Brasil**.
- **Conclusão:** Diante disso, o Conselho Digital defende uma abordagem baseada na **compreensão técnica da inteligência artificial**, na **proporcionalidade regulatória** e no **interesse público**. O fair use ou uso justo maximiza a inovação, mas reduz a escolha (além de não ser um dispositivo do direito brasileiro); o opt-in maximiza o controle formal, mas inviabiliza o treinamento de sistemas de IA; **o opt-out é o ponto intermediário funcional**. Quando implementado de forma razoável e **legível por máquina**, o opt-out protege direitos, preserva a diversidade linguística e cultural, mantém a inovação e estabelece bases sólidas para uma governança democrática do treinamento de sistemas de inteligência artificial. Além disso, é a melhor solução enquanto um dispositivo legal específico não é aprovado.

Para pensar: Imagine o seguinte, caso o opt-in obrigatório fosse implementado, e desenvolvedores de programas de IA em

geral teriam de pedir, por exemplo, autorização para o titular de cada artigo, imagens, layouts ou base de dados disponíveis de forma aberta na internet para desenvolver sua pesquisa. Considerando a quantidade de páginas acessíveis de forma online – estimada em bilhões – previsões restritivas como o opt-in poderiam inviabilizar por completo o treinamento de modelos de IA no Brasil.

MODELOS DE TRATAMENTO REGULATÓRIO DO TREINAMENTO DE MÁQUINAS

Modelo regulatório	Como funciona	Pontos positivos	Pontos negativos	Impactos estruturais
Uso justo¹	Fair use é uma doutrina do copyright que permite certos usos de obras protegidas sem autorização do autor, desde que o uso seja considerado justo após uma análise do caso concreto. O treinamento de sistemas de IA é realizado independente de exclusão pelos titulares de conteúdo.	<ul style="list-style-type: none"> ■ Máxima viabilidade técnica ■ Baixo custo de transação ■ Forte estímulo à inovação, pesquisa e concorrência ■ Facilita entrada de novos atores 	<ul style="list-style-type: none"> ■ Pode gerar judicialização ■ Dificulta negociação e compensação ■ Gera resistência social e política ■ Não é aplicável ao direito brasileiro 	<ul style="list-style-type: none"> ■ Ecossistema altamente inovador ■ Pouca governança formal
Opt-out	O treinamento é permitido como regra, mas detentores de direitos autorais podem se excluir de forma expressa.	<ul style="list-style-type: none"> ■ Compatível com a escala do treinamento de IA ■ Preserva o direito de escolha ■ Mantém pluralidade e concorrência ■ Cria base objetiva para negociação 	<ul style="list-style-type: none"> ■ Exige implementação técnica clara ■ Depende de padrões comuns (ex.: legível por máquina) ■ Pode ser utilizado para tolher modelos já treinados, inviabilizando a prática 	<ul style="list-style-type: none"> ■ Equilíbrio entre inovação e controle ■ Infraestrutura para governança dinâmica ■ Viabiliza mercados e acordos voluntários
Opt-in	O treinamento só é permitido mediante autorização prévia e expressa de cada titular.	<ul style="list-style-type: none"> ■ Máximo controle formal para titulares ■ Simplicidade conceitual 	<ul style="list-style-type: none"> ■ Tecnicamente inviável em escala² ■ Altíssimo custo de transação ■ Favorece concentração econômica ■ Exclui universidades e desenvolvedores menores 	<ul style="list-style-type: none"> ■ Redução drástica da inovação ■ Dependência de poucos atores globais ■ Enfraquecimento da soberania tecnológica
Permissão Legal Expressa	É possível que a legislação preveja de forma expressa a não aplicabilidade da proteção por direitos autorais em relação ao treinamento de modelos de IA.	<ul style="list-style-type: none"> ■ Maior segurança jurídica para desenvolvedores, que podem acessar livremente qualquer conteúdo disponível online (desde que de forma legítima). 	Ausência de gestão por parte dos titulares de direitos autorais, impossibilidade de oposição.	<ul style="list-style-type: none"> ■ Acelera a inovação, ao retirar entraves para a atividade de desenvolvimento/treinamento de modelos.

¹ Do inglês *fair use*, fruto da tradição anglo-saxã, não pode ser confundida com a “regra dos três passos” prevista pela Convenção de Berna (1886), da qual o Brasil é signatário. Essa Convenção prevê em seu art. 9(2) que: As exceções só são permitidas quando: (i) Se aplicam a casos especiais; (ii) Não prejudiquem a exploração normal da obra; (iii) Não causem prejuízo injustificado aos interesses legítimos do autor.

² Ordem de grandeza de dados para treinar IA: ML tradicional: 10³–10⁵ exemplos. Deep Learning: 10⁵–10⁷ exemplos. LLMs médios: 10⁹–10¹¹ tokens. LLMs grandes: 10¹²–10¹³ tokens